MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2017/2018

DSP5018 –STATISTICS AND PROBABILITY

(A111, A114-116, B271-273, B275-283)

01 JUNE 2018

09.00 am – 11.00 am

(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of 4 pages excluding cover page and appendix.

2. Attempt **ALL FOUR (4)** questions. The marks distribution for each question is given.

3. Write your answers in the answer booklet provided.

4. Key formulae are given in the Appendix.

Question 1

a) Fill in the blanks with the choices given in the box. (5 marks)

(please write the answer in your answer sheet)

i. Statistics is the field of study implies collecting, presenting, analyzing and interpreting _____ in order to make decision.

ii. Inferential statistics consists of method for drawing and _____ the reliability of conclusions about a population based on information obtained from a sample of the population.

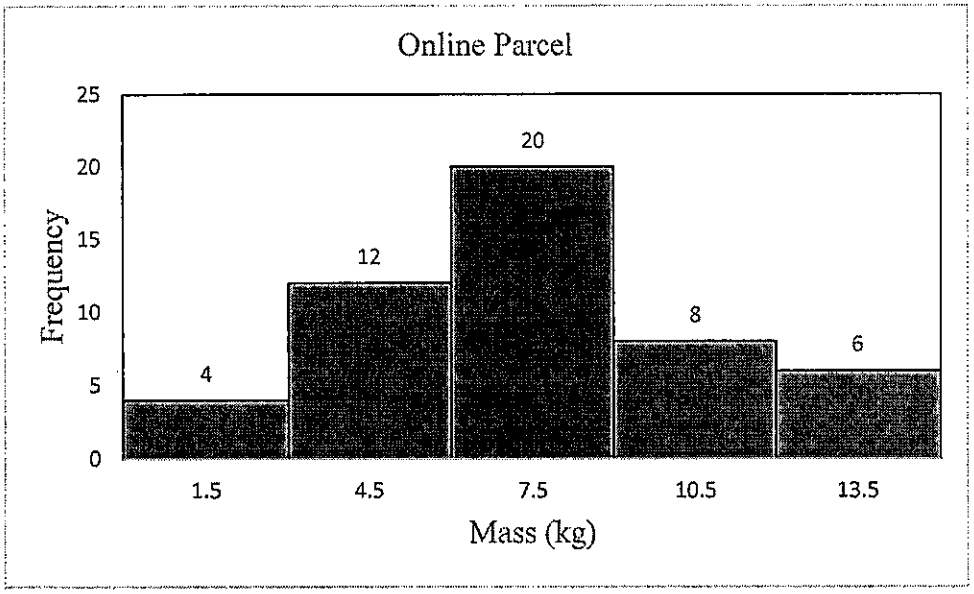
iii. Two major types of statistics are _____ and inferential.

iv. Sample is the part of the _____ from which information is obtained.

v. Discrete variable is a quantitative variable whose possible _____ can be listed.

measuring	population	descriptive
values	data	mod

b) Diagram below shows a histogram of the masses of online parcels received by a company.



i. From the data shown in the diagram above, complete the following table. (4 marks)

(please reconstruct the table in your answer sheet)

Masses (kg)	Frequency	Relative Frequency	Midpoint
0⩽3			

Continued...

Using your answer in (i), calculate

ii. sample mean. (3 marks)

iii. sample standard deviation. (4 marks)

c) Given 5 Number Summary: Minimum = 3, Q_1 = 6, Median = 12, Q_3 = 16 and Maximum = 21.

i. Sketch a boxplot. (3 marks)

ii. Calculate the lower limit and upper limit. (6 marks)

[TOTAL 25 MARKS]

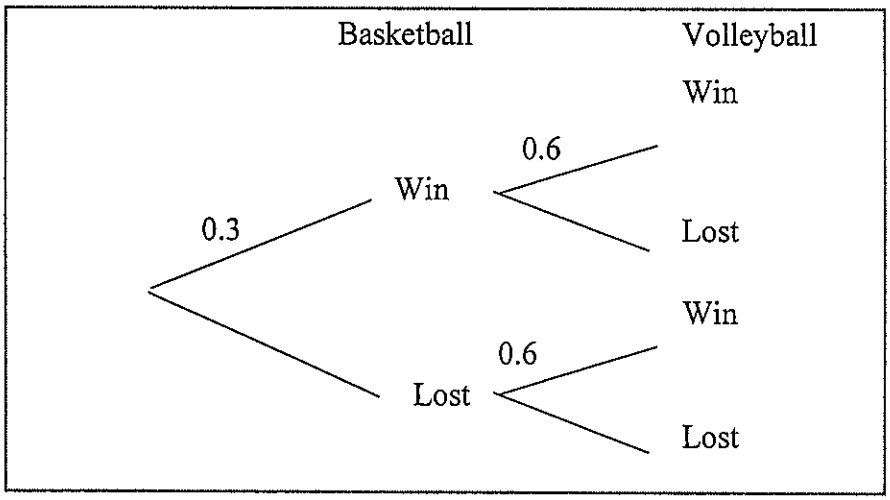
Question 2

a) Identify each statement below TRUE / FALSE. (5 marks)

(please write the answer in your answer sheet provided)

No	Statement	TRUE / FALSE
i.	The set of all possible outcomes of an experiment is called an event.	
ii.	In some experiments it is helpful to list the elements by tree diagram or Venn diagram.	
iii.	The probability of an event that cannot occur is 1.	
iv.	The probability that event B occurs given that event A occurs is called a conditional probability.	
v.	An event that must occur is called a certain event.	

b) Halim is going to play one basketball match and one volleyball match. Suppose he has a 30% and 60% chances of winning a basketball match and volleyball match respectively as shown below.



Continued...

i. From the tree diagram above, complete the table below. (6 marks)
(please reconstruct the table in your answer sheet)

Outcome	Probability
Win , Win	
	0.12

ii. What is the probability that Halim will win either one of the match? (2 marks)

c) A box contains 5 energy bars and 3 yogurts. Tan eats 2 of them, by randomly choosing.
i. List two possible elements of the choices that Tan eat. (2 marks)
ii. What is the probability of choosing 1 energy bar and 1 yogurt? (4 marks)

d) JUMBO supermarket gather information of their membership card according to the gender and age as below:

Age	Gender		Total
	Male	Female	
18 – 25	12	7	19
26 – 50	45	62	107
> 50	43	11	54
Total	100	80	180

If a member is randomly chosen by the supermarket, calculate the probability of getting
i. a member age not more than 50 years old. (2 marks)
ii. a male member given that the member age more than 50 years old. (2 marks)
iii. Are event Male and Female mutually exclusive or not? Give reason to your answer. (2 marks)

[TOTAL 25 MARKS]

Question 3

a) The water quality in Melaka was tested for heavy metal contamination. The average heavy metal concentration from a sample of 29 different locations is 3 grams per milliliter with a standard deviation of 0.5. Construct the 95% confidence intervals for the mean heavy metal concentration. (4 marks)

b) The average weekly usage of electricity in MMU was measured for 35 weeks, and 2 of the biggest buildings were compared.

Building A	Building B
$n_1 = 35$	$n_2 = 35$
$\overline{X_1} = 75.3$	$\overline{X_2} = 71.5$
$\sigma_1 = 10.6113$	$\sigma_2 = 6.5491$

Continued...

Based on the results, can we conclude that the average weekly usage of electricity in the two buildings are different at 5% significance level? (10 marks)

c) For the upcoming World Cup 2018, a survey of 1500 football fans in Malaysia found that 1170 of them preferred to watch the game at mamak compared to watching at home. At 1% significance level, can we conclude that more than 75% of football fans in Malaysia prefer to watch World Cup 2018 at mamak? (11 marks)

[TOTAL 25 MARKS]

Question 4

A random sample of eight drivers insured with a company and having similar auto insurance policies was selected. The following table lists their driving experiences, x (in years) and monthly auto insurance premiums, y (in RM).

Driving Experience, x (Years)	Monthly Auto Insurance Premium, y (RM)
5	180
2	201
12	154
9	169
15	148
6	176
25	120
16	136

a) Given $\sum y^2 = 210854, \sum x^2 = 1396, \sum xy = 13123$. Compute the value of SS_{xx}, SS_{yy} and SS_{xy}. (3 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (3 marks)

c) Compute the regression equation $\hat{y} = \beta_0 + \beta_1 x$. (4 marks)

d) If a person has experienced on driving for 20 years, what is his or her monthly auto insurance premium can we expect? (2 marks)

e) Find the sum of squares of *SST*, *SSR* and *SSE*. (3 marks)

f) Calculate the standard error of the estimate, s_e. (2 marks)

g) At 5% significance level, do the data provide sufficient evidence to conclude that the driving experience and monthly auto insurance premium are negatively linearly correlated? (8 marks)

[TOTAL 25 MARKS]

End of page.

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^{\,2} f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$ where m : class midpoint f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

ESTIMATION

Margin of Error, *MoE*	
$MoE = (\pm 1.96)(s_{\bar{x}})$ where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$	$MoE = (\pm 1.96)(\sigma_{\bar{x}})$ where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

The $(1-\alpha)100\%$ confidence interval for population mean	
$\bar{x} \pm \left(z_{\frac{\alpha}{2}}\right)(\sigma_{\bar{x}})$ where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm \left(z_{\frac{\alpha}{2}}\right)(s_{\bar{x}})$ where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$	$\bar{x} \pm \left(t_{\frac{\alpha}{2},n-1}\right)(s_{\bar{x}})$ where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ degree of freedom, $df = n - 1$

The $(1-\alpha)100\%$ confidence interval for population proportion
$\hat{p} \pm \left(z_{\frac{\alpha}{2}} \right) \left(s_{\hat{p}} \right)$ where $\hat{p} = \frac{x}{n}$, $s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$, $\hat{q} = 1 - \hat{p}$

z_α				
$z_{0.10}$	$z_{0.05}$	$z_{0.025}$	$z_{0.01}$	$z_{0.005}$
1.282	1.645	1.960	2.326	2.576

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ where $\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$

REGRESSION

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \frac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$
- **Correlation coefficient,** $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$